EXHIBIT C

Biochem. J. (1986) 240, 1–12 (Printed in Great Britain)

REVIEW ARTICLE

The purification of eukaryotic polypeptides synthesized in *Escherichia coli*

Fiona A. O. MARSTON
Protein Biochemistry Department, Celltech Ltd., 244–250 Bath Road, Slough, Berks. SL1 4DY, U.K.

INTRODUCTION

Over the last 13 years, manipulation of DNA *in vitro* has developed from the transfer of genetic information between prokaryotic organisms (Cohen *et al.*, 1973) to a technology which facilitates efficient and controlled production of proteins in foreign hosts. A significant feature of these developments is the ability to express eukaryotic genes in prokaryotes such as *Escherichia coli* (Harris, 1983; Wetzel & Goeddel, 1983). The supply of many eukaryotic polypeptides which have potential clinical or industrial use is often limited by their low natural availability. Gene cloning and expression in *E. coli* can provide a more abundant source of these polypeptides.

The mode of gene expression affects the location of the proteins produced. The proteins may either be located in the cytoplasm of *E. coli* or secreted through the cell membrane. Eukaryotic genes cloned in frame with synthetic or bacterial nucleic acid sequences can be expressed as hybrid products in the cell cytoplasm. Transcription, from bacterial promoters, and translation, yield fusion proteins which include bacterial or synthetic polypeptide sequences in addition to the eukaryotic polypeptide. An alternative approach which locates proteins in the cytoplasm is direct expression, where bacterial promoters and terminators are used in the transcription of the foreign gene alone. In *E. coli* an ATG, or occasionally a GTG, sequence must precede the gene coding sequence, for translation initiation. Thus the primary products of translation possess an *N*-terminal methionine residue. *E. coli* possesses enzymes which catalyse the efficient removal of the methionine residues from natural proteins when required, but these enzymes do not work with the same efficiency on recombinant polypeptides and therefore directly expressed proteins may possess an unnatural *N*-terminal methionine residue. Finally, gene sequences which include a leader or signal sequence cloned in frame with the eukaryotic genes, when transcribed and translated can direct secretion of the eukaryotic polypeptides through the bacterial cell membrane.

From the increasing number of reports of eukaryotic polypeptide synthesis in *E. coli* it is clear that the mode of expression affects not only the efficiency of production, but the nature of the polypeptide product itself. In general, recombinant polypeptides accumulate to higher levels of total cell protein when expressed intracellularly than when secreted, but many of the polypeptide products located in the cytoplasm are insoluble and aggregated. The consequent isolation and purification techniques required are the subject of this Review.

INTRACELLULAR EXPRESSION

Genes expressed directly or as fusion proteins in the cytoplasm of *E. coli* characteristically accumulate to levels ranging up to 25% of total cell protein (Table 1). However, in the majority of cases, the expressed proteins are in an insoluble form (Harris, 1983). This was not expected, as the authentic proteins are naturally produced in soluble forms. The appearance of inclusion bodies in *E. coli* in parallel with the accumulation of proinsulin, insulin A chain or insulin B chain (Williams *et al.*, 1982) was the first indication that the insoluble proteins might accumulate in a discrete form. By isolating inclusion bodies from cells expressing prochymosin it was demonstrated that these inclusions were indeed predominantly composed of recombinant protein (Marston *et al.*, 1984). A number of eukaryotic polypeptides expressed in *E. coli* directly, e.g. bovine growth hormone, salmon growth hormone, IFN-β, IFN-γ, IL-2 and Protein C, or as fusion proteins, e.g. proinsulin, myoglobin and β-globin, have now been shown to exist as aggregates or inclusion bodies (see Table 1 for references).

In the phase contrast microscope, inclusion bodies are seen to be highly refractile, while transmission electron microscopy reveals them as amorphous aggregates not enclosed or in contact with a distinct membrane (Schoemaker *et al.*, 1985; Schoner *et al.*, 1985). Electron micrographs show isolated inclusions as spherical in shape (Fig. 1*a*), although this may not be the conformation *in vivo*, because of the preparation procedures used before microscopy. Contaminating material can be seen, associated with the isolated inclusion bodies (Figs. 1*a* and 1*b*).

There is no direct evidence to indicate why eukaryotic polypeptides are sequestered into inclusion bodies in *E. coli*. The accumulation of abnormal *E. coli* proteins in intracellular granules was demonstrated some years ago (Prouty & Goldberg, 1972; Prouty *et al.*, 1975). However, normal *E. coli* proteins synthesized to high levels using recombinant DNA techniques can also accumulate in insoluble forms (Gribskov & Burgess, 1983; Botterman & Zabeau, 1985) and as inclusion bodies (Cheng, 1983). It is therefore not simply a response to 'foreign' proteins. One interpretation is that

Abbreviations used: AIDS, acquired immune deficiency syndrome; BPV, bovine papilloma virus; CAT, chloramphenicol acetyltransferase; EGF, epidermal growth factor; FMDV, foot and mouth disease virus; HBV, hepatitis B virus; HSV, herpes simplex virus; IFN, interferon; IGF, insulin-like growth factor; IL, interleukin; TNF, tumour necrosis factor.

Table 1. Properties of some eukaryotic polypeptides located in the cytoplasm of *E coli*

Polypeptide	M_r ($\times 10^{-3}$) of authentic non-glycosylated polypeptide	Mode of expression	Expression (%)	Location on cell lysis†	No. of cysteine residues	References
Somatostatin	1.5	Fusion	< 0.05	Pellet	2	Itakura *et al.* (1977)
Insulin A chain	2.0	Fusion	20	Pellet	4	Goeddel *et al.* (1979*b*)
Insulin B chain	3.0	Fusion	20	Pellet	2	Goeddel *et al.* (1979*b*)
Calcitonin	3.5	Fusion	17	Pellet	2	Bennett *et al.* (1984)
β-Endorphin	3.9	Fusion	5	Pellet	0	Shine *et al.* (1980)
Urogastrone	7.5	Fusion	NE‡	Pellet	6	Sassenfeld & Brewer (1984)
T4 *RegA*	14.6	Fusion	NE	Pellet	1	Adari *et al.* (1985)
β-Globin	16	Fusion	5–10	Pellet	2	Nagai *et al.* (1985)
Myoglobin	17	Fusion	10	Pellet	0	Varadarajan *et al.* (1985)
Bovine growth hormone	22	Fusion	5	Pellet	2	Seeburg *et al.* (1978)
Human growth hormone	22	Fusion	5	Pellet	2	Szoka *et al.* (1986)
*α_1-Antitrypsin	45	Fusion	15	(Supernatant)	1	Courtney *et al.* (1984)
*Complement C5a	8.3	Direct	0.007	(Supernatant)	7	Mandecki *et al.* (1985)
*Interleukin-2	12–17	Direct	10	(Supernatant)	3	Devos *et al.* (1983)
*AIDS peptide 121	15	Direct	5–10	Pellet	2	Chang *et al.* (1985)
Human TNF	17	Direct	15	Supernatant	2	Pennica *et al.* (1984)
Murine TNF	17	Direct	24	Supernatant	2	Pennica *et al.* (1985)
*Interferon γ	18	Direct	25	(Supernatant)	2	Simons *et al.* (1984)
*Human lymphotoxin	18	Direct	NE	Supernatant	0	Gray *et al.* (1984)
Interferon α	19.4	Direct	NE	Supernatant	5	Staehelin *et al.* (1981) ; Wetzel *et al.* (1981)
*Interferon β	22–26	Direct	15	Pellet	3	Whitehorn *et al.* (1985)
Growth hormone (bovine)	22	Direct	NE	Pellet	2	George *et al.* (1985)
Growth hormone (human)	22	Direct	NE	Supernatant	2	Goeddel *et al.* (1979*a*)
Growth hormone (salmon)	22	Direct	15	Pellet	2	Sekine *et al.* (1985)
κ-Light chain (IgG)	24	Direct	0.5	Pellet	5	Cabilly *et al.* (1984)
AIDS p24 gag	24	Direct	NE	Supernatant	0	Dowbenko *et al.* (1985)
Apolipoprotein E	34.2	Direct	1	Pellet	1	Vogel *et al.* (1985)
Calf prochymosin	43	Direct	8	Pellet	6	Marston *et al.* (1984)
*γ-Heavy chain (IgG)	49	Direct	3	Pellet	13	Cabilly *et al.* (1984)
*Protein C	50	Direct	(25% of insoluble protein)	Pellet	24	Hoskins *et al.* (1985)
Triosephosphate isomerase	53	Direct	0.3	Supernatant	4	Straus & Gilbert (1985)
Urokinase	54	Direct	NE	Pellet	24	Winkler *et al.* (1985)
MMLV reverse transcriptase	80	Direct	20	(Supernatant)	8	Kotewicz *et al.* (1985)

* Polypeptides that are naturally glycosylated *in vivo*.
† Parentheses indicate that only a proportion of the activity is located in the supernatant.
‡ NE, not estimated.

proteins aggregate when they are synthesized at such a high rate that the cells' degradation systems become saturated (Prouty *et al.*, 1975). This cannot explain why prochymosin expressed at low levels is also insoluble (Schoemaker *et al.*, 1985). What is evident, however, is that stringent chemical conditions, as described below, are required to solubilize recombinant proteins from inclusion bodies. Formation is therefore not just a precipitation phenomenon resulting from the accumulation of a high concentration of protein. Precipitation may be an initial effect, but at some stage ionic, hydrophobic or covalent interactions form between the protein molecules.

When recombinant polypeptides are insoluble, purification does not just involve the application of chromatographic separation techniques, the objective being also to recover active, soluble protein. Fig. 2 illustrates diagrammatically procedures that can be used to solubilize aggregated recombinant polypeptides. For directly expressed proteins, (Fig. 2*a*), the first stage is to isolate the inclusion bodies. Then denaturants are used to unfold the polypeptides and finally conditions are adjusted to allow the polypeptides to refold correctly. For fusion proteins it may in addition be necessary to cleave the fusion to isolate the recombinant polypeptide (Fig. 2*b*, stage 3). A varying degree of purification can be achieved during these procedures, and once the polypeptides are solubilized conventional chromatographic techniques can be applied.

The fact that recombinant polypeptides aggregate in a discrete form is useful for the purposes of purification. As first noted for abnormal *E. coli* proteins (Prouty *et al.*, 1975), inclusion bodies are dense and sediment readily with low speed centrifugation. Speeds as low as 500 *g* are

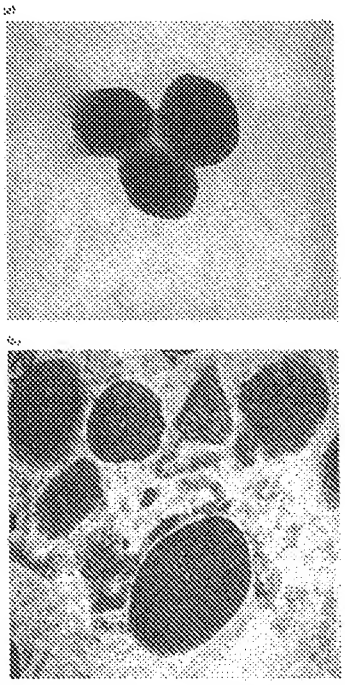

Fig. 1. Electron micrographs of prochymosin inclusion bodies isolated from *E. coli*

Sample preparation: inclusion bodies were isolated from *E. coli* HB101 pCT 70 as described in Marston *et al.* (1984). The inclusion bodies were fixed first in 2.5% glutaraldehyde then in 1% OsO_4. After dehydration through a series of alcohols, the inclusion bodies were embedded in Spurr's resin. Sections 60 nm thick were cut using an LKB Ultratome III. Sections were stained using uranyl acetate (30 min, 60 °C) and lead citrate (10 min, room temperature). (*a*) Transmission electron micrograph of water-washed isolated inclusion bodies in suspension. Magnification × 63000. (*b*) Transmission electron micrograph of a thin section through a pellet of inclusion bodies. Magnification × 63000. (Provided by and reproduced with the permission of Richard Sugrue, Ray Newsam and the University of Kent Electron Microscopy Unit, Canterbury, U.K.).

reported to sediment recombinant inclusion bodies (Olsen, 1985) but values of 5000–12000 *g* are more generally used (Marston *et al.*, 1984; George *et al.*, 1985; Schoner *et al.*, 1985). Under these conditions, the inclusion bodies sediment more rapidly than the bulk of the cell debris and purification is achieved. However, contaminating proteins do co-purify with the inclusions, as illustrated for prochymosin (Fig. 3). Detergent (Marston *et al.*, 1984) or urea (Koths *et al.*, 1985; Schoner *et al.*, 1985) can be used to solubilize the microbial proteins preferentially, leaving the recombinant proteins between 30% and 90% pure. The usefulness of this procedure is underlined by the fact that techniques have been developed to enhance aggregation. When a proportion of the human growth hormone expressed in *E. coli* was found to partition in the soluble fraction, methods normally used to kill cells, such as heat treatment (60–80 °C), acid treatment, or phenol plus toluene, were used to increase recovery of the recombinant protein in an aggregated form (Olsen, 1985).

Fusion proteins

The intracellular accumulation of a eukaryotic polypeptide expressed directly in *E. coli* may be limited because the protein is recognized as foreign and degraded. This is particularly apparent with small polypeptides (Wetzel & Goeddel, 1983). Expression levels can be improved by linking the eukaryotic gene with a bacterial gene and producing a fused protein product, as demonstrated for somatostatin (Itakura *et al.*, 1977), insulin (Goeddel *et al.*, 1979*b*) and β-endorphin (Shine *et al.*, 1980). Fusion proteins can be expressed to levels of up to 26% of total cell protein (Table 1), but if the bacterial gene constitutes a large proportion of the fusion, then the amount of eukaryotic product will be small. There are many examples of fusion proteins which accumulate in an insoluble form (Table 1) despite the fact that the eukaryotic sequence is a small proportion of the total protein sequence; typical examples being β-endorphin (31 amino acids) fused to β-galactosidase (Shine *et al.*, 1980) and calcitonin-Gly (32 amino acids) fused to CAT (Bennett *et al.*, 1984).

Another fusion method is one in which multiple copies of the gene are linked in tandem (Shen, 1984). Expression levels were increased when two or more linked proinsulin genes were expressed directly or in conjunction with a small part of the *N*-terminus of β-galactosidase. The stability of the two products was similar, but the yield of the β-galactosidase fusion was 3-fold greater than that of the directly expressed product. Effects at the level of transcription, translation or on mRNA stability were suggested. Enhanced expression of proinsulin in *E. coli* by the *N*-terminal addition of short homo-oligopeptides was described in a recent report (Sung *et al.*, 1986). Seven of the 20 oligomers studied were particularly effective, but the mechanism by which accumulation was affected was not established.

Fusion proteins may be purified and used without further modification. The isolation of inclusion bodies can be utilized as a purification step for insoluble proteins (Kleid *et al.*, 1981; Pilacinski *et al.*, 1984; Cabradilla *et al.*, 1986). Further purification methods used include preparative electrophoresis (Kleid *et al.*, 1981), immunopurification (Liu *et al.*, 1984), gel filtration and ion-exchange chromatography (Cabridilla *et al.*, 1986). In order to maintain the proteins in a soluble form for column chromatography, detergents or denaturants may be included in the buffers (Liu *et al.*, 1984; Cabradilla *et al.*, 1986). Purified intact fusion proteins have been used in the development of vaccines for FMDV (VP1), BPV and cholera toxin (Kleid *et al.*, 1981; Pilancinski *et al.*, 1984; Jacob *et al.*, 1985), in the development of diagnostic kits for the AIDS retrovirus (Cabradilla *et al.*, 1986) and to demonstrate biological activity of the F_c portion of IgE (Liu *et al.*, 1984).

When the eukaryotic polypeptide is required in isolation from the fusion protein the strategy used is to place a cleavage site between the *C*-terminus of the prokaryotic sequence and the *N*-terminus of the eukaryotic coding sequence; Table 2 lists some typical examples. A schematic illustration of a typical isolation and cleavage protocol for fusion proteins is given in Fig. 2. Inclusion bodies are isolated and denatured before cleavage. Chemical cleavage can be effected using CNBr, which cleaves on the *C*-terminal side of the methionine residues. This method was used in the production of

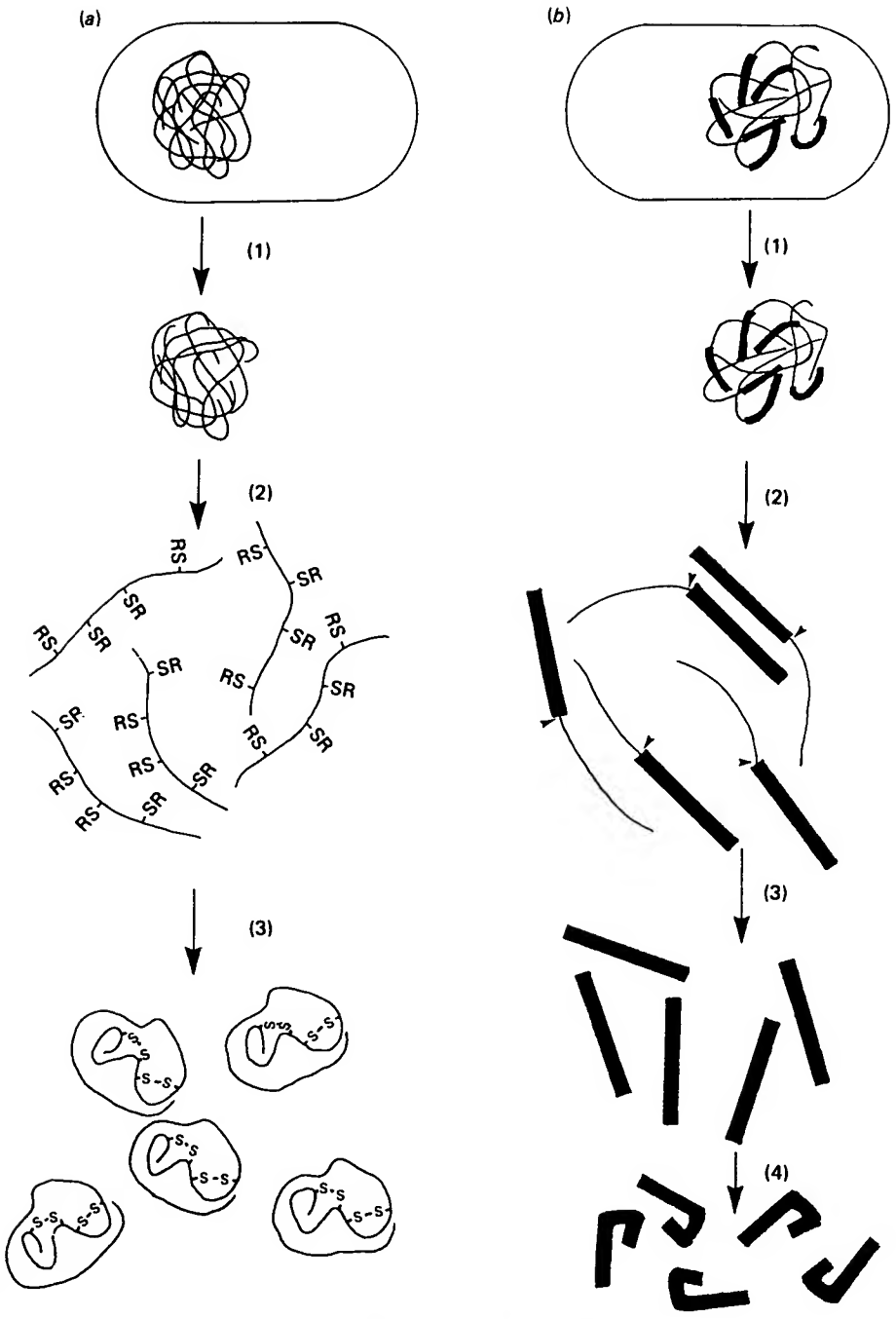

Fig. 2. Diagram of stages in the recovery of active proteins from the cytoplasm of *E. coli*

(*a*) Direct expression. Stage 1, cell lysis and isolation of inclusion bodies. Stage 2, denaturation. Stage 3, refolding. The example illustrated is a disulphide-containing protein. R represents either –H or $-SO_3^-$. (*b*) Fusion proteins. Stages 1 and 2 as for (*a*). Stage 3, cleavage (▼) to release the recombinant polypeptide. Stage 4, refolding; ■ represents the eukaryotic portion of the fusion protein.

β-galactosidase–insulin A chain and β-galactosidase–insulin B chain fusions (Goeddel *et al.*, 1979*b*). Both products were insoluble and centrifugation was therefore used as a purification step. Denaturation using guanidinium chloride in the presence of β-mercaptoethanol was necessary before CNBr cleavage. The individual chains were purified further by using ion-exchange chromatography, reverse phase h.p.l.c. and gel filtration. *S*-Sulphonated derivatives of the chains mixed and re-oxidized yielded biologically active insulin. The tandem linked pro-insulin genes expressed by Shen (1984) were each separated by the sequence -Arg-Arg-Asn-Ser-Met-. The intervening sequences were cleaved after the methionine residues by CNBr treatment. This method is limited in its use, since most proteins are likely to contain methionine residues. Also, it is necessary for the polypeptides to be acid-stable as cleavage is performed in 70% formic acid.

A unique cleavage site, not present in the coding sequence of the recombinant gene, is the ideal;

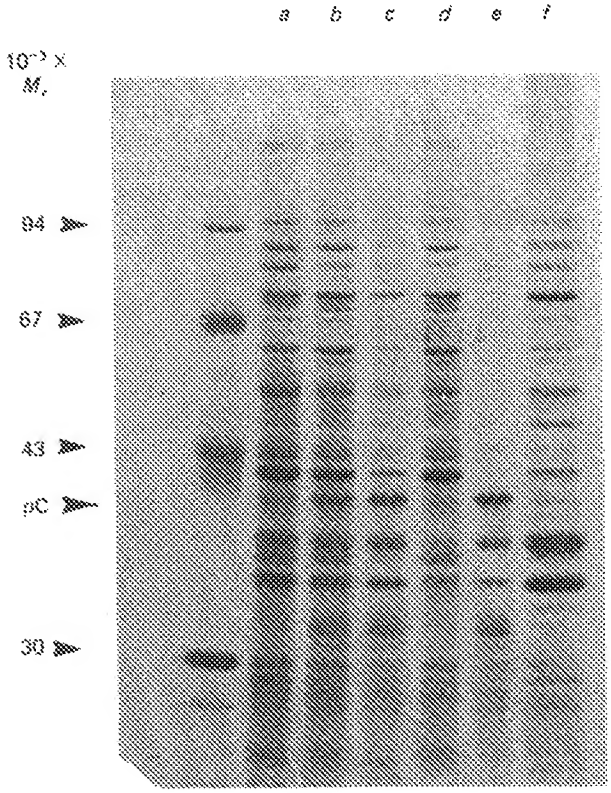

Fig. 3. Coomassie Blue stained SDS/polyacrylamide-gel profile of total *E. coli* proteins and isolated inclusions

Lane *a*, total cell proteins from *E. coli* HB101/pCT54 (control); lane *b*, total cell proteins from *E. coli* HB101/pCT70 (prochymosin producer). Lanes *c*–*f* were all from HB101/pCT70. Lane *c*, lysis pellet; lane *d*, lysis supernatant; lane *e*, Triton X-100/EDTA-washed pellet; lane *f*, Triton X-100/EDTA-wash supernatant. M_r markers are indicated on the left. pC, prochymosin.

otherwise, the recombinant product will also be cleaved. Use of a lysine link in a TrpE–EGF fusion was possible because EGF contains no lysine residues (Allen *et al.*, 1985). Isolated inclusion bodies were solubilized in 8 M-urea, and after dilution, the EGF was released from the fusion protein by digestion with endoproteinase Lys C. Gel filtration and ion-exchange chromatography yielded EGF 80–90% pure. Batch variation in the activity of endoproteinase Lys C rendered this cleavage procedure irreproducible. TrpE–EGF fusion proteins containing acid-labile (-Asp-Pro-), factor Xa- and collagenase-sensitive linking regions were also described, but little if any EGF was released from any of these fusions using the appropriate cleavage conditions. If the cleavage site is not unique, then the internal lysis sites can be protected. This was possible with a β-galactosidase–β-endorphin fusion (Shine *et al.*, 1980) where internal lysine residues were reversibly blocked by citraconylation and trypsin was used to cleave after the-Lys-Arg- residues immediately before the *N*-terminus. After deblocking of the lysines, a biologically active product was recovered. However, this blocking and deblocking approach is not generally satisfactory.

An insoluble TrpE–bovine growth hormone fusion was constructed with an acid-labile Asp-Pro cleavage site

Table 2. Cleavage sites engineered into fusion proteins in *E. coli*

indicates the peptide bond cleaved.

Sequence recognized	Cleavage effector	Reference
-Asp▼Pro-	Acid pH	Szoka *et al.* (1986)
-Met▼Xaa-	CNBr	Goeddel *et al.* (1979*b*)
-Arg▼Xaa- or -Lys▼Xaa-	Trypsin	Shine *et al.* (1980)
-Ile-Glu-Gly-Arg▼Xaa	Factor X_a	Nagai *et al.* (1985)
-Pro-Xaa▼Gly-Pro-Yaa▼	Collagenase	Lee & Ullrich (1984)
-Arg▼Xaa	Clostripain	Bennett *et al.* (1984)

(Szoka *et al.*, 1986). Low pH treatment was performed in the presence of guanidinium chloride. The bovine growth hormone released bound to growth hormone receptors *in vitro*. However, this cleavage regime leaves proteins with an uncharacteristic *N*-terminal proline residue, and the acid conditions used may cause amide loss from asparagine residues.

The samples described above illustrate the fact that insoluble fusion proteins may have to be cleaved in the presence of denaturant. This is a consideration when an enzymically cleavable site is engineered. Clostripain cleavage of CAT–calcitonin is possible, as the enzyme is stable in up to 6 M-urea (Bennett *et al.*, 1984). Carboxypeptidase B, used to cleave urogastrone–polyarginine, is stable in up to 5 M-urea (Sassonfeld & Brewer, 1984).

The longer the recognized cleavage site, the less likely it is to be found in the coding sequence of recombinant protein. A specific example is the cleavage sequence recognized by collagenase (Table 2) which has been introduced into TrpE–IGF and TrpE–EGF fusion proteins (Lee & Ullrich, 1984). A tetrapeptide sequence, -Ile-Glu-Gly-Arg-, precedes the cleavage sites of factor Xa. Both human β-globin (Nagai *et al.*, 1985) and human myoglobin (Varadarajan *et al.*, 1985) have been fused via this sequence to a λcII protein fragment. The resulting fusion proteins were insoluble and inclusion bodies were isolated. Triton-washed inclusions were solubilized in urea. β-Globin was purified further by ion-exchange chromatography and gel filtration. Denaturant was removed before cleavage with factor Xa. Refolded β-globin was reconstituted with haem and α-globin. The oxygen-binding properties of the recombinant haemoglobin were essentially the same as those of authentic human haemoglobin (Nagai *et al.*, 1985).

In the protocol for myoglobin, inclusion bodies were isolated and washed. The intact fusion proteins were reconstituted with haem in the presence of urea. This allowed cleavage by trypsin in place of factor Xa, because the holoenzyme is resistant to proteolytic attack. Dialysis against Tris buffer, pH 8.0, was used to remove the trypsin. Ion-exchange chromatography and gel filtration were used to purify the protein further. Using the trypsin cleavage process, myoglobin has been produced on the gram scale (Varadarajan *et al.*, 1985).

If a proenzyme such as prochymosin is expressed as a fusion, insertion of a specific cleavage site is not required. At acid pH, prochymosin is autocatalytically processed to the chymosin and the fusion protein is removed with the pro-peptide (Nishimori *et al.*, 1984).

There are intracellular fusion proteins which are soluble. Examples include β-galactosidase fusions with HBV pre-S2 region (Offensperger *et al.*, 1985) and cholera toxin CTP3 (Jacob *et al.*, 1985). λcII protein sequences fused to α_1-antitrypsin (Courtney *et al.*, 1984) and λN protein sequences fused to HSV–thymidine kinase (Waldman *et al.*, 1983). The latter is an interesting example of a fusion protein in which there was no specific cleavage site. However, the fusion was correctly processed by an *E. coli* proteinase during cell lysis, liberating active thymidine kinase. α_1-Antitrypsin was not completely soluble, and approx. 60% of the fusion protein was insoluble and inactive (Courtney *et al.*, 1984).

Fusion proteins can be constructed to facilitate purification. A bacterial sequence can be selected that codes for a polypeptide which can be isolated by affinity chromatography. The β-galactosidase–HBV pre-S2 fusion synthesized in *E. coli* was purified using *p*-aminophenyl-β-D-thiogalactoside–Sepharose. In a single step, the protein was purified more than 30-fold to >90% homogeneity (Offensperger *et al.*, 1985). Substrate affinity chromatography can also be used to purify polypeptides fused to CAT (Bennett *et al.*, 1984), an appropriate substrate being acetyl-CoA. The latter approach is limited to soluble fusion products. Typically, cleavage sites are needed to allow the production of free polypeptide.

A different approach has been taken in designing the polyarginine purification fusion (Sassenfeld & Brewer, 1984). A synthetic gene sequence coding for polyarginine was fused to the 3′ end of the urogastrone gene. Cation-exchange chromatography was used to purify the positively charged recombinant fusion protein. This is an effective purification step, as most bacterial proteins are acidic, and are negatively charged at the pH of 5.5 which was used. The polyarginine sequence can be cleaved using carboxypeptidase B to yield free urogastrone. At this stage, a second cation-exchange step retains the contaminating proteins while the urogastrone flows straight through. The urogastrone–polyarginine was insoluble, so the purification and proteolytic cleavage with carboxypeptidase B was performed in the presence of urea. It was suggested that fusion of this peptide to other recombinant polypeptides would allow the same purification methodology to apply (Brewer & Sassenfeld, 1985). However, the recombinant polypeptide to which the polyarginine is fused may interact with, and restrict the availability of the polyarginine to bind to, a matrix. Another important factor is the efficiency with which the polyarginine can be removed to leave the correct *C*-terminus.

Direct expression

The major purification problem for directly expressed products is the development of techniques to release them from aggregates into stable active and soluble forms. The first step is to solubilize the proteins and conditions are used which, *in vitro*, will denature native proteins. The solubilization agents used include: (*a*) 5–8 M-guanidinium chloride (insulin A and B chains, bovine growth hormone and urokinase). (*b*) 6–8 M-urea (IgG heavy and light chains, prochymosin, IFN-γ, and salmon growth hormone), (*c*) detergents (IFN-β and IL-2), (*d*) alkaline pH (> 9.0) (prochymosin and chicken growth hormone), and (*e*) organic solvents (bacteriophage T4 *regA* protein) (see Table 1 for references).

Solubilization is therefore achieved by disrupting non-covalent hydrogen bonds, ionic or hydrophobic interactions and unfolding the polypeptides. The effectiveness of a particular solvent is likely to differ between proteins; being dependent on the nature of the polypeptides themselves. There are a number of variables in the protocols published, including pH, temperature, time and ionic environment. Another consideration is the ratio of denaturant to protein (Marston *et al.*, 1984), which can affect the efficiency of the overall process.

Solubilized proteins can be purified in the presence of denaturant. Proteins can be subjected to ion-exchange or gel filtration, if urea, alkaline pH or non-ionic detergents have been used (Gill *et al.*, 1985; George *et al.*, 1985). Since guanidinium chloride is charged, gel filtration but not ion-exchange chromatography may be used (Builder & Ogez, 1985; Gill *et al.*, 1985). Purification has also been achieved by using high-speed centrifugation to remove contaminating microbial proteins (Builder & Ogez, 1985) and by preferential organic extraction of recombinant proteins into butan-2-ol or 2-methylbutan-1-ol (Konrad & Lin, 1984; Koths *et al.*, 1985).

Having disrupted the aggregated polypeptides, conditions must be adjusted to allow refolding. The recovery of high yields of activity depends on the polypeptides refolding with the formation of the correct intramolecular interactions, including disulphide bonds. Dialysis has been used, successfully, to remove denaturant and generate soluble, active bovine growth hormone and urokinase (George *et al.*, 1985; Winkler *et al.*, 1985), although for urokinase it was necessary to maintain the protein concentration at or below a critical level (Winkler *et al.*, 1985). In contrast, dialysis of IFN-γ generated both active monomeric and inactive aggregated protein (Arakawa *et al.*, 1985). The low yields of active prochymosin obtained using dialysis (McCaman *et al.*, 1985) were much improved by using dilution to reduce the urea concentration (Marston *et al.*, 1984). These data are consistent with the results of studies of authentic protein folding *in vitro* which show that protein concentration is an important parameter in optimizing the yield of correctly folded protein (London *et al.*, 1974; Mozhaev & Martinek, 1982). The dilution must be such that intramolecular interactions occur in preference to intermolecular interactions.

The pH used during refolding can also affect the yield obtained. Exposure to alkaline pH (> 9.0) without the use of urea or guanidinium chloride has been used to unfold proteins, renaturation being effected by reduction of pH (Lowe *et al.*, 1984). However, dilution from urea into buffers at alkaline pH was used to produce active salmon growth hormone (Sekine *et al.*, 1985) and prochymosin (Marston *et al.*, 1984). In many of the examples quoted by Olsen (1985), alkaline pH was used at some stage of the refolding processes. pH may be critical for the formation of correct disulphide bond formation, since thiol–disulphide interchange proceeds more rapidly at alkaline pH (Freedman & Hillson, 1980).

In a series of three patents, an unfolding and refolding

process was described based on the use of 'strong' denaturants followed by the use of a 'weak' denaturant, urea (Builder & Ogez, 1985; Olsen, 1985; Olsen & Pai, 1985). Guanidinium chloride was classed as a 'strong' denaturant. Unusually, Triton, SDS and chaotropic salts such as thiocyanate were classified in this category. It was suggested that, after unfolding, transfer into low concentrations of urea permits refolding of the proteins into forms approximating their native state. This process was used to refold FMDV capsid protein, porcine growth hormone, prochymosin (prorennin) and urokinase.

Only the existence of non-covalent interactions in inclusion bodies has been discussed so far. In certain protocols, used successfully to solubilize recombinant polypeptides, thiol reagents were used in conjunction with the denaturants (Cabilly *et al.*, 1984; George *et al.*, 1985; Koths *et al.*, 1985; Vogel *et al.*, 1985; Chang *et al.*, 1985; Winkler *et al.*, 1985). If the inclusion of thiol reagents is essential for the recovery of activity, covalent disulphide bonds may exist in the aggregates Many of the eukaryotic polypeptides insoluble when expressed in the *E. coli* cytoplasm are normally secreted by their natural cells and contain disulphide bonds which form between cysteine residues during the process of secretion. In common with a number of bacteria *E. coli* maintains its cytoplasm in a reduced state (Fahey *et al.*, 1977). In limited studies, bacterial cytoplasmic proteins were found to have a low cysteine content and to contain few disulphide bonds (Pollock & Richmond, 1962; Fahey *et al.*, 1977). Taking into consideration the high level at which eukaryotic polypeptides are synthesized, and the potential number of disulphide bonds, it would seem unlikely that these bonds form *in vivo* in the cytoplasm of *E. coli*. It is probable that they form on exposure to air, during cell lysis, with many of the disulphide bond arrangements being incorrect. However, when prochymosin was expressed directly in *E. coli*, a small proportion of oxidized monomer was shown to exist in intact cells. It was suggested that these disulphide bonds may have formed in an oxidizing microenvironment within the inclusion bodies (Schoemaker *et al.*, 1985).

If intramolecular or intermolecular disulphide bonds in inclusion bodies are incorrect, then disruption will be an essential component in the recovery of activity. Thiol reagents have been included in denaturation buffers and maintained in all the buffers used up to the refolding stage (Olsen, 1985). Alternatively, reversible *S*-sulphonation (Katsoyannis *et al.*, 1966) has been used (Cabilly *et al.*, 1984; Olsen, 1985). Further formation of disulphide bonds was prevented until the redox conditions were altered to allow thiol–disulphide interchange during the refolding step.

It should be noted that reduction may not be essential for solubilization or refolding, as exemplified by prochymosin (Marston *et al.*, 1984) and growth hormone (George *et al.*, 1985; Gill *et al.*, 1985). There is no evidence that incorrect disulphide bonds cause aggregation. There is limited evidence that some disulphide bonds form *in vivo* (Schoemaker *et al.*, 1985) but they are clearly a feature of the isolated inclusions bodies.

The disruption of disulphide bonds may be a consideration in the denaturation stage, and conversely the formation of correct disulphide bonds may be critical during refolding. Growth hormones contain two disulphide bonds per molecule. It is interesting to note that while thiol reagents may be included during the unfolding step for bovine growth hormone (George *et al.*, 1985) none are used in the refolding step for bovine and chicken growth hormones (George *et al.*, 1985; Gill *et al.*, 1985). Less than 2% of the refolded growth hormone preparations were in an aggregated form. Thiol reagents were omitted entirely in unfolding and refolding salmon growth hormone (Sekine *et al.*, 1985). Thus these growth hormones appear to form native disulphide bonds in the absence of exogenous thiol reagents. A similar observation was made for prochymosin (Marston *et al.*, 1984) and it was suggested that thiol–disulphide interchange was promoted by free thiol groups in the recombinant polypeptide.

When exogenous addition of thiols is required to promote correct disulphide bond formation by thiol–disulphide interchange, a mixture of reduced and oxidized reagents, such as glutathione, may be included in the refolding buffer. This may follow denaturation in the absence (Winkler *et al.*, 1985) or presence of thiol reagent (Olsen, 1985) or after *S*-sulphonation (Cabilly *et al.*, 1984; Olsen, 1985). For IL-2, oxidation was promoted during refolding using iodosobenzoic acid (Koths *et al.*, 1985).

The recovery of biological activity has been demonstrated for many of the refolded polypeptides described in this section. Examples include immunoglobulins (Boss *et al.*, 1984; Cabilly *et al.*, 1984), IL-2 (Liang *et al.*, 1985; Malkovsky *et al.*, 1985), bovine, chicken and salmon growth hormones (George *et al.*, 1985; Gill *et al.*, 1985; Sekine *et al.*, 1985), prochymosin (Green *et al.*, 1985) and urokinase (Winkler *et al.*, 1985). However, with few exceptions (Boss *et al.*, 1984; Cabilly *et al.*, 1984; Marston *et al.*, 1984), quantification of the efficiency with which the proteins are refolded is not provided.

As described earlier, insoluble recombinant polypeptides can be purified to different extents before or during refolding. Using either gel filtration or ion-exchange chromatography during refolding, growth hormones > 90% pure were obtained (Gill *et al.*, 1985). Dialysis was used as a final step in these processes, and precipitation was observed. If contaminating microbial proteins precipitated preferentially, then this may have enhanced the final purity of the growth hormones.

If the isolation and washing of inclusion bodies yields high purity recombinant protein, only a limited number of chromatographic steps may subsequently be required to yield pure protein. For prochymosin, a single ion-exchange column yielded protein > 90% pure. Acid-activation of the zymogen produces chymosin > 99% pure. The yield of active refolded material was > 40% (Marston *et al.*, 1984). Refolded urokinase was purified using a five-step process (Winkler *et al.*, 1985). Two of the column steps were affinity purification using benzamidine–Sepharose and the high yield from these columns was significant in generating the overall yield of 32% of refolded protein.

Some effort has been directed at eliminating disulphide bond formation in order to facilitate the recovery of active, soluble protein from *E. coli*. IFN-β and IL-2 each contain three cysteine residues, and a single disulphide bond is thought to exist in each native protein. When expressed in *E. coli*, IFN-β and IL-2 are aggregated and insoluble. Mutant proteins or 'muteins' of IFN-β and IL-2 were synthesized in *E. coli* after site-directed mutagenesis, in which one cysteine residue was either

deleted or replaced by a serine residue (Mark *et al.*, 1983; Koths *et al.*, 1985). The polypeptides still formed insoluble aggregates and required denaturation and refolding. The mutant proteins were biologically active once refolded. In contrast, deletion of the three *N*-terminal residues (Cys-Tyr-Cys) of IFN-γ was reported to result in a more soluble *E. coli* protein (Allet, 1985). Some characterization studies have been performed using the refolded protein (Hsu & Arakawa, 1985), but details of the purification process, which is in preparation, are required to allow the effect of this deletion to be interpreted. Another example of a modification which affected solubility is provided by Moloney murine leukaemia virus (MMLV) reverse transcriptase. This enzyme was predominantly insoluble when expressed in *E. coli*. Insertion of a termination codon earlier in the 3′ sequence deleted a hydrophobic *C*-terminal region and resulted in the expression of a totally soluble protein (Kotewicz *et al.*, 1985). This result implicates hydrophobic interactions in aggregation, consistent with results obtained when cloned viral glycoprotein genes are expressed in *E. coli* (Harris, 1984).

Only insoluble, directly expressed polypeptides have been discussed. There are several examples of polypeptides expressed in soluble, active forms in *E. coli*. For certain polypeptides only a small proportion of the expressed protein is soluble, but this facilitates characterization of the recombinant product. Receptor binding of human complement fragment C5a (Mandecki *et al.*, 1985) and biological activity of FMDV proteinase (Klump *et al.*, 1984) were demonstrated using unpurified lysis supernatants. Pure retroviral p24 gag protein was purified from *E. coli* in a single immunopurification step (Dowbenko *et al.*, 1985) while a five-step process was used to produce human growth hormone (Olsen *et al.*, 1981). Other examples of soluble proteins are human IFN-α (Staehelin *et al.*, 1981; De Maeyer *et al.*, 1982) bovine IFN-α (Grosfeld *et al.*, 1985), chicken triosephosphate isomerase (Straus & Gilbert, 1985), human lymphotoxin (Gray *et al.*, 1984), human TNF (Pennica *et al.*, 1984) and murine TNF (Pennica *et al.*, 1985).

There are no obvious common features to explain why these proteins are soluble. Expression levels varied from $< 1\%$–25% of total cell protein. Some, but not all, of the authentic proteins are naturally glycosylated. The fact that none of these proteins contain large numbers of disulphides may be significant. However, there are several examples of polypeptides containing one or two disulphides which are totally insoluble (Table 1).

Authentic proteins can be unfolded and refolded *in vitro* with little loss of activity. Yields are critically dependent on a number of parameters, particularly protein concentration, and for disulphide bond-containing proteins, pH. Such studies have led to the conclusion that the amino acid sequence of polypeptides contains the information required for folding (Anfinsen, 1973). Why is it, therefore, that eukaryotic polypeptides synthesized in *E. coli* and having the correct amino acid sequence fail to fold correctly? Insolubility does not result just because the proteins are expressed at a high percentage of total cell protein, as was observed for overexpressed *E. coli* proteins. There are examples of eukaryotic polypeptides expressed to levels of 1% or less which are insoluble (Table 1).

The mechanism by which proteins fold *in vivo* is still unknown. From studies *in vitro* it is evident that the amino acid sequence of each protein contains the information required for folding, but it is not apparent which residues specify the folding information. Another consideration is what influence, if any, the chemical environment within the cell has on protein folding. In the absence of such information it is only possible to speculate why some eukaryotic polypeptides fail to fold correctly in *E. coli*.

For proteins which contain disulphide bonds, formation of these bonds may be an essential component of the folding process. The inability to form disulphide bonds in the reduced environment of the *E. coli* cytoplasm may thus prevent folding. However this may not be the only reason why these proteins do not fold correctly, and certainly does not explain why eukaryotic proteins which lack disulphide bonds can also fail to fold in *E. coli*.

While there is little data available about the chemical environment in *E. coli* and eukaryotic cells *in vivo*, there are some obvious differences, notably the lack of subcellular compartments in *E. coli*. Therefore it is possible for the nascent recombinant polypeptide chains to come into contact with low-M_r components or macromolecules in the *E. coli* cytoplasm which are isolated in organelles in their natural cell environment. Interaction with these components could interfere with or inhibit protein folding.

The structure of the folding eukaryotic polypeptides may also be important if differences in structure result in the proteins being recognized as foreign. It may be for this reason that the proteins are sequestered into aggregates as discussed earlier. In this respect it is interesting that chicken triosephosphate isomerase, a cytoplasmic protein of M_r 53000, which is soluble in *E. coli*, has a structure which is highly conserved across prokaryotes and eukaryotes (Straus & Gilbert, 1985).

Our knowledge of the mechanisms by which proteins fold *in vivo* is limited and, as highlighted in a recent review (King, 1986), solving the protein folding problem is important for development in this area of biotechnology.

SECRETED PROTEINS

E. coli possesses two cell membranes, the outer membrane and the cytoplasmic membrane, which are separated by the periplasmic space. Proteins located in the periplasmic space or the outer membrane are synthesized in the cytoplasm and exported through the cytoplasmic membrane (Michaelis & Beckwith, 1982). The proteins are synthesized as precursors, with an *N*-terminal signal sequence which may be cleaved during secretion. Signal sequences are an absolute requirement for export (Silhavy *et al.*, 1983) but there may also be information in the mature protein sequence which affects its localization (Tommassen *et al.*, 1983; Ghrayeb *et al.*, 1984; Freundl *et al.*, 1985).

Very few *E. coli* proteins are secreted into the extracellular medium, across the outer membrane (Muller *et al.*, 1983). While a signal sequence is required for periplasmic secretion there is no evidence that such sequences are required for secretion across the outer cell membrane. Enterotoxins contain signal sequences which are cleaved during secretion (Dallas & Falkow, 1980), while haemolysin is released extracellularly without cleavage of the signal peptide. The mechanism by which

proteins cross the outer membrane is unclear but appears to be more complex than for periplasmic secretion.

Expression via secretion offers several advantages over intracellular expression, as discussed in a recent review comparing *E. coli*, yeast and *Bacillus subtilis* as secretion systems for foreign proteins (Nicaud *et al.*, 1986). If the signal sequence is correctly processed, the *N*-terminus of the recombinant protein will be identical to the authentic product. Secretion of proteins into the periplasm can prevent the degradation of the polypeptides; for example, proinsulin located in the periplasm was 10-fold more stable than when located in the cytoplasm (Talmadge & Gilbert, 1982). Disulphide bond formation in *E. coli* has been shown to occur simultaneously with secretion (Pollitt & Zalkin, 1983). Folded eukaryotic products with correct disulphide bonds can therefore be produced, as demonstrated for proinsulin (Emerick *et al.*, 1985) and EGF (Oka *et al.*, 1985).

Periplasmic secretion

It was suggested that *E. coli* cannot process eukaryotic signal peptides efficiently (Devos *et al.*, 1983) but more recent evidence is contradictory. The eukaryotic signal of IgG light chain both initiated secretion and was correctly processed (Zemel-Dreasen & Zamir, 1984) but when the prokaryotic β-lactamase signal sequence was placed in front of the IgG light chain signal sequence only the bacterial signal peptide was processed during secretion. The natural leader sequences of urokinase (Jacobs *et al.*, 1985) and human growth hormone (Gray *et al.*, 1985) were both processed by *E. coli*, though secretion was only demonstrated with the latter. Prokaryotic signal sequences from β-lactamase (Villa-Komaroff *et al.*, 1978), OmpA protein (Ghrayeb *et al.*, 1984) and alkaline phosphatase (*phoA*) (Oka *et al.*, 1985) have been used successfully to secrete eukaryotic polypeptides into the periplasm of *E. coli*.

To assess whether proteins have been secreted, *E. coli* must be converted to spheroplasts to release the contents of the periplasm. A β-lactamase proinsulin fusion was found to secrete over 90% of the proinsulin synthesized into the periplasmic space (Chan *et al.*, 1981). The level of expression was 0.01% of total cell protein. Using genetic and physiological manipulations, the level of expression was improved to 0.5% of total protein (Emerick *et al.*, 1984). Human EGF has also been secreted into the periplasm, using the *phoA* signal peptide. The strains expressing the highest level of the polypeptide secreted 1–2 mg/l of culture (approx. 0.01–0.02% of total cell protein (Oka *et al.*, 1985).

In both the examples cited above, the signal sequences were correctly processed. The expression levels are apparently low, when compared with the levels achieved by direct expression in the cytoplasm. However, the polypeptides are soluble and biologically active. If the M_r of the polypeptide secreted is small, than even at expression levels of 0.02%, a significant number of molecules are formed (1.8×10^5 molecules/cell for EGF). It has been suggested that for insulin an expression level of 1% would be adequate for commercial production (Emerick *et al.*, 1984).

Periplasmic proteins are normally only 4% of total cell protein (Nossal & Heppel, 1966); therefore, less extensive purification of recombinant proteins is required than for proteins located in the cytoplasm. EGF has been purified to apparent homogeneity in a two-step process; gel filtration followed by reverse-phase h.p.l.c. (Oka *et al.*, 1985).

Secretion of eukaryotic polypeptides can prove toxic to *E. coli*, causing cell lysis. This was observed with OmpA–FMDV VP1 fusions (Henning *et al.*, 1983) and β-lactamase–proinsulin fusions (Brosius, 1984). It was suggested that the proteins may become anchored in the cytoplasmic membrane, blocking export completely. This was previously observed for bacterial *lamB–lacZ* fusions (Emr *et al.*, 1980).

Extracellular secretion

Since the mechanism by which proteins are secreted into the extracellular medium by *E. coli* is poorly understood, limited effort has been directed at developing this expression system. In an attempt to secrete the eukaryotic polypeptide, β-endorphin, a fusion with the signal peptide and part of the *N*-terminal sequence of the OmpF protein was constructed (Nagahari *et al.*, 1985). The β-endorphin was secreted into the culture medium, with the signal peptide correctly processed. Secretion was dependent on the presence of the signal sequence. The level of β-endorphin in the medium was estimated to be 1–2 mg/litre of culture, which represented > 99% of the total β-endorphin in the periplasm and medium. Negligible amounts were detected in the cytoplasm. In contrast, the percentage of the periplasmic enzymes β-lactamase and alkaline phosphatase detected in the medium was approx. 14% and 11% respectively, indicating a certain level of periplasmic leakage or cell lysis. The mechanism of secretion was not established, but the *N*-terminal region of the OmpF protein, included in the fusion, may play a role in secretion.

The β-endorphin polypeptides were purified from the medium by gel filtration through Sephadex G-10 and reverse phase h.p.l.c. No indication of starting purity was given, but column profiles (monitored at 230 nm) showed a number of peaks in addition to those identified as β-endorphin. Sequence analysis and amino acid composition of the purified polypeptides revealed *C*-terminal heterogeneity, which could result from premature termination of transcription or proteolytic degradation.

Another eukaryotic polypeptide secreted at significant levels into the medium by *E. coli* was A α-fibrinogen (Lord, 1985). This polypeptide has a predicted M_r of 67000. When expressed as a fusion with the β-lactamase signal, 4 mg/l were located in total cell lysates, and 13 mg/l in the culture medium, but comparative levels of cytoplasmic or periplasmic enzymes were not presented. However, if levels of cell lysis are low, then these results demonstrate that large polypeptides can be secreted into the medium by *E. coli*. Preliminary results indicated a correct *N*-terminus but some heterogeneity at the *C*-terminus.

CHARACTERISTICS OF THE *E. COLI* EXPRESSION SYSTEM

There is concern over the use of *E. coli* as a production system because of the existence of endotoxins or lipopolysaccharide in the cell wall. In fact, lipopolysaccharide contamination could be introduced into any manufactured drug or biological regardless of the organism of origin. Lipopolysaccharide is pyrogenic and therefore must be removed to yield a safe product. There are several chromatographic techniques which can be

used to remove lipopolysaccharide. Matrices which have been used include polymixin B–Sepharose, histamine–Sepharose, substrate-analogue–Sepharose, DEAE-Sepharose (Sofer, 1984). Many of these procedures are commonly used in protein purification.

As a result of being synthesized in *E. coli* recombinant eukaryotic polypeptides can differ from their authentic counterparts. The possibility of an additional methionine residue at the *N*-terminus of directly expressed polypeptides was mentioned earlier. There are also a number of eukaryotic post-translational modifications of polypeptides which *E. coli* does not perform. These include glycosylation, acetylation and amidation. *E. coli* has a requirement for an ATG (or GTG) codon at the 5′ end of a gene, as translation is initiated by *N*-formylmethionine. The methionine is deformylated in *E. coli* during synthesis (Adams, 1968), but is not necessarily cleaved. Thus directly expressed polypeptides may possess an uncharacteristic *N*-terminal methionine. *E. coli* does possess an aminopeptidase with broad substrate specificity, but the efficiency of methionine removal from recombinant polypeptides is variable. The *N*-terminal methionine residues were completely removed from IFN-β (Stebbing *et al.*, 1982) and bovine growth hormone (George *et al.*, 1985), whilst there was differential processing of IFN-α (Staehelin *et al.*, 1981) and human growth hormone (Olsen *et al.*, 1981). With an *N*-terminal sequence of Met-Ala-Pro-, IL-2 was completely unprocessed (Liang *et al.*, 1985). However, when a deletion mutant, lacking the three *N*-terminal residues of the authentic protein (Ala-Pro-Thr-) was expressed, the *N*-terminus was Met-Ser-, and all the methionine was removed. Sequence effects were also observed with bovine growth hormone, Met-Phe- being much less efficiently processed than Met-Ala- (Seeburg *et al.*, 1983). One explanation is that residues near to the *N*-terminus cause steric hindrance and thus affect the efficiency of methionine removal (Liang *et al.*, 1985). The advantage of using fusion proteins or secretion is that processing yields the correct *N*-terminal residue. It is possible for the presence of an *N*-terminal methionine to have no effect on biological activity, as demonstrated for human growth hormone and IL-2 (Olsen *et al.*, 1981; Liang *et al.*, 1985).

E. coli does not possess the enzymes necessary for glycosylating polypeptides. There are however documented examples of recombinant proteins synthesized in *E. coli* which are biologically active despite the fact that they are not glycosylated. These include IFN-β, IFN-γ (Edge & Camble, 1984) and IL-2 (Liang *et al.*, 1985). These results not only demonstrate that carbohydrate residues are not essential for biological activity, but also that when the expressed product is insoluble they are not essential for refolding. Acetylation and amidation are two other post-translational processes which are often essential for the biological activity of some eukaryotic polypeptides. It has therefore been necessary to develop systems *in vitro* to perform these reactions on recombinant polypeptides isolated from *E. coli*. Acetylation *in vitro* of desacetylthymosin (Wetzel *et al.*, 1981) and amidation of calcitonin (S. K. Rhind, personal communication) have been achieved.

Recombinant polypeptides which have residues missing at the *N*-terminus or C-terminus may be produced in *E. coli* together with the intact molecules. These may arise as a result of late starts or premature termination in transcription. Alternatively, proteolysis may have occurred *in vivo* or *in vitro*. The product of the *lon* gene in *E. coli*, protease La, has an important role in the degradation of abnormal proteins. The fact that *lon* gene transcription increases with the synthesis of recombinant polypeptides (Goff & Goldberg, 1985) suggests that the use of *lon*$^-$ strains might decrease degradation. Results have varied from no effect (Emerick *et al.*, 1984) to increased expression of 150-fold (Boss *et al.*, 1984). Full consideration of this subject is beyond the scope of this Review, but it is worth noting that there can be considerable variation in the accumulation of recombinant proteins in different *E. coli* strains (Schoemaker *et al.*, 1985). The extent to which the intact molecules must be purified from incorrect length molecules will depend upon the use for which the recombinant polypeptides are being produced.

CONCLUDING REMARKS

Because the knowledge of *E. coli* genetics was well advanced, it became the focus for the development of recombinant DNA techniques and historically, for the same reasons, *E. coli* was selected as *the* organism in which to express eukaryotic polypeptides. From the information described in this Review, it is clear that there are both advantages and disadvantages to the use of this expression system.

Using intracellular expression, gram amounts of polypeptide can be produced per litre of fermentation, although the products are commonly insoluble. While secretion of proteins from *E. coli* obviates the insolubility phenomenon, until recently expressed yields have been low. The production of milligram amounts per litre of β-endorphin (Nagahari *et al.*, 1985) and A α-fibrinogen (Lord, 1985) show that *E. coli* has potential as a secretion system. For further development this secretion mechanism must be elucidated.

Denaturation and refolding has been used to produce a number of biologically active eukaryotic polypeptides from *E. coli*. These include proteins with complex multidomain structures such as urokinase (Winkler *et al.*, 1985) and tissue-type plasminogen activator (Harris *et al.*, 1986). There are currently insufficient data to allow detailed discussion of the efficiency with which many of the recombinant polypeptides refold. However, the levels of biological activity recovered in the examples described were clearly adequate for analytical studies.

The fact that recombinant polypeptides could become modified during the unfolding and refolding process must be considered. Exposure to high concentrations of denaturants such as urea and high pH must be minimized because of the possibility of chemical modification. To characterize the proteins, the assignment of disulphide bonds for recombinant and authentic products has been compared (Kohr *et al.*, 1982). Gross conformation can be probed by c.d., while detailed structural information can be provided by n.m.r. and X-ray crystallography. In this respect the results for refolded *E. coli* β-globins are encouraging. X-ray diffraction analysis of 0.28 nm (2.8 Å) resolution revealed only slight electron density differences between authentic and mutant recombinant products. This difference could be explained by the Cys→Ser change engineered by site-directed mutagenesis (Nagai *et al.*, 1985).

In conclusion, *E. coli* can be used to produce larger amounts of eukaryotic polypeptides than can be isolated

from natural cell sources. The removal of lipopolysaccharide was highlighted earlier as an important aim of purification. If adequate precautions are taken, such as the use of specific chromatographic steps coupled with the use of pyrogen-free water and maintenance of equipment in a pyrogen-free state, then lipopolysaccharide can be reduced to an acceptably low level. This is borne out by the fact that there are therapeutic recombinant polypeptides from *E. coli* in use. IL-2, IFN-α, IFN-β, IFN-γ and TNF are among the recombinant polypeptides currently in clinical trials, while insulin and human growth hormone are examples of *E. coli*-derived polypeptides which are now in clinical use.

I am grateful to my colleagues Peter Lowe, Tim Harris, Spencer Emtage, Rick Lilley and Martyn Robinson for their constructive criticisms. I would like to thank Margaret Turner for her patience in typing this manuscript. Thanks also to Peter Dumbell and Alan Lyons for assistance in the preparation of Figures. The data of Sofer (1984) was used by permission of Bio/Technology © 1984.

REFERENCES

Adams, J. M. (1968) J. Mol. Biol. **33**, 571–589
Adari, H Y., Rose, K., Williams, K. R., Konisberg, W. H., Lin, T.-C. & Spicer, E. K. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 1901–1905
Allen, G., Paynter, C. A. & Winther, M. D. (1985) J. Cell Sci. Suppl. **3**, 29–38
Allet, B. (1985) World Pat. WO85/04186
Anfinsen, C. B. (1973) Science **181**, 223–230
Arakawa, T., Alton, N. K. & Hsu, Y.-R. (1985) J. Biol. Chem. **260**, 14435–14439
Bennett, A. D., Rhind, S. K., Lowe, P. A. & Hentschel, C. C. G. (1984) Eur. Pat. Appl. 0131363
Boss, M. A., Kenten, J. H., Wood, C. R. & Emtage, J. S. (1984) Nucleic Acids Res. **12**, 3791–3806
Botterman, J. & Zabeau, M. (1985) Gene **37**, 229–239
Brewer, S. J. & Sassenfeld, H. M. (1985) Trends Biotechnol. **3**, 119–122
Brosius, J. (1984) Gene **27**, 161–172
Builder, S. E. & Ogez, J. R. (1985) U.S. Pat 4511502
Cabradila, C. D., Groopman, J. E., Lanigen, J., Renz, M., Lasky, R. A. & Capon, J. A. (1986) Bio/Technology **4**, 128–133
Cabilly, S., Riggs, A. D., Pande, H., Shirely, J. E., Holmes, W. E., Rey, M., Perry, L. J., Wetzel, R. & Heyneker, H. L. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 3273–3277
Chan, S. J., Weiss, J., Konrad, M., White, T., Bahl, C., Yu, S. D., Mark, D. & Steiner, D. F. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 5401–5405
Chang, T. W., Kato, I., McKinney, S., Chanda, P., Barone, A. D., Wong-Staal, F., Callo, R. C. & Chang, N. T. (1985) Bio/Technology **3**, 905–909
Cheng, Y. S. E. (1983) Biochem. Biophys. Res. Commun. **111**, 104–111
Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 3240–3244
Courtney, M., Buchwalder, A., Tessier, L.-H., Jaye, M., Benavente, A., Ballard, A., Kohli, V., Lathe, R., Tolstoshev, P. & Lecocq, J.-P. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 669–673
Dallas, W. S. & Falkow, S. (1980) Nature (London) **288**, 499–501
De Maeyer, E., Skup, D., Prosad, K. S. N., De Maeyer-Guignard, J., Williams, B., Meacock, P., Sharpe, G., Proli, D., Hennam, J., Scuch, W. & Atherton, K. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 4256–4259
Devos, R. Plaetinck, G., Cheroutre, H., Simons, G., Degrave, W., Tavernier, J., Renalt, E. & Fiers, W. (1983) Nucleic Acids Res. **11**, 4307–4323
Dowbenko, D. J., Bell, J. R., Benton, C. V., Groopman, J. E., Nguyon, H., Vetterlein, D., Capen, D. J. & Laskey, L. A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7748–7752
Edge, M. D. & Camble, R. (1984) Biotechnol. Genet. Eng. Rev. **2**, 215–252
Emerick, A. W., Bertolani, B. L., Ben-Bassat, A., White, T. J. & Konrad, M. W. (1984) Bio/Technology **2**, 165–168
Emr, S. D., Hedgpeth, J., Clement, J.-M., Silhavy, T. J. & Hofnung, M. (1980) Nature (London) **285**, 82–85
Fahey, R. C., Hunt, J. S. & Windham, G. C. (1977) J. Mol. Evol. **10**, 155–160
Freedman, R. B. & Hillson, D. A. (1980) in The Enzymology of Post-Translational Modification of Proteins (Freedman, R. B. & Hawkins, H. C., eds.), vol. 1, pp. 165–166, Academic Press, New York
Freundl, R., Schwarz, H., Klose, M., Movva, N. R. & Henning, U. (1985) EMBO J. **4**, 3593–3598
George, H. J., L'Italien, J. J., Pilancinski, W. P., Glassman, D. L. & Krzyzek, R. A. (1985) DNA **4**, 273–281
Gill, J. A., Sumpter, J. P., Donaldson, E. M., Dye, H. M., Souza, L, Berg, T., Wypch, J. & Langley, K. (1985) Bio/Technology **3**, 643– 646
Goeddel, D. V., Heyneker, H. L., Hozumi, T., Arentzon, R., Itakura, K., Yansura, D. G., Ross, M. J., Miozzari, G., Crea, R. & Seeburg, P. (1979*a*) Nature (London) **281**, 544–548
Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. & Riggs, A. D. (1979*b*) Proc. Natl. Acad. Sci. U.S.A. **76**, 106–110
Ghrayeb, J., Kimura, H., Takashara, M., Hsiong, H., Masui, Y. & Inovye, M. (1983) EMBO J. **3**, 2437–2442
Goff, S. A. & Golberg, A. L. (1985) Cell **41**, 587–595
Goldberg, M. E. (1985) Trends Biochem. Sci. **10**, 388–391
Gray, G. L., Baldridge, J. S., McKeown, K. S., Heyneker, H. L. & Chang, C. N. (1985) Gene **39**, 247–254
Gray, P. W., Aggarwal, B. B., Benton, C. V., Bringman, T. S., Henzel, W. S., Jarnett, J. A., Leung, D. W., Moffat, B., Ng, P., Svederesky, L. P., Palladino, M. A. & Medwin, G. E. (1984) Nature (London) **312**, 721–724
Green, M. L., Angal, S., Lowe, P. A. & Marston, F. A. O. (1985) J. Dairy Res. **52**, 281–286
Gribskov, M. & Burgess, R. R. (1983) Gene **26**, 109–118
Grosfeld, H., Cohen, S., Shafferman, A. & Velan, B. (1985) Brit. U.K. Pat. Appl. 2157697
Halling, S. M. & Smith, S. (1985) Bio/Technology **3**, 715–720
Harris, T. J. R. (1983) in Genetic Engineering (Williamson, R., ed.), vol. 4, pp. 127–185, Academic Press, London
Harris, T. J. R. (1984) in Immune Intervention (Roitt, I. M., ed.), vol. 1, pp. 57–92, Academic Press, London
Harris, T. J. R., Patel, T., Marston, F. A. O., Little, S., Emtage, J. S., Opdenakker, G., Volckaert, G., Rombauts, W., Billiau, A. & De Somer, P. (1986) Mol. Biol. Med., in the press
Henning, U., Cole, S. T., Bremer, E., Hindernach, I. & Schaller, H. (1983) Eur. J. Biochem. **136**, 233–240
Holmes, W. E., Pennica, D., Blaber, M., Rey, M. W., Guenzler, W. A., Steffers, G. J. & Heyneker, H. L. (1985) Bio/Technology **3**, 923–929
Hoskins, J. A., Schoner, B. E., Belogoye, R. M. & Long, G. L. (1985) Thromb. Haemostasis **54**, 167
Hsu, Y.-R. & Arakawa, T. (1985) Biochemistry **24**, 7959–7963
Itakura, K., Hiroso, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. & Boyer, H. W. (1977) Science **198**, 1056–1063
Jacob, C. O., Leitne, M., Zamir, A., Salomon, D. & Arnon, R. (1985) EMBO J. **12**, 3339–3343
Jacobs, P., Cravador, A., Loriau, R., Brockly, F., Colau, B., Churchana, X., Van Elsen, A., Herzog, A. & Bollen, A. (1985) DNA **4**, 139–146
Katsoyannis, P. G., Tometsko, A. & Zalut, C. (1966) J. Am. Chem. Soc. **88**, 166–167
King, J. (1986) Bio/Technology **4**, 297–303

Kleid, P. G., Yanasura, D., Small, B., Dowbenko, D., Moore, D., Grubman, M., McKercher, P., Morgan, D. O., Robertson, B. H. & Bachroch, H. L. (1981) Science **214**, 1125–1129
Klump, W., Marquadt, O. & Hofschneider, P. H. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 3351–3355
Kohr, W. J., Keck, R. & Harkins, R. N. (1982) Anal. Biochem. **122**, 348–359
Konrad, M. W. & Lin, L. S. (1984) U.S. Pat. 4450103
Kotewicz, M. L., D'Alessio, J. M., Driftmier, K. M., Blodgett, K. P. & Gerard, G. F. (1985) Gene **35**, 249–258
Koths, K. E., Thompson, J. W., Kunitani, M., Wilson, K. & Hanisch, W. H. (1985) Eur. Pat. Appl. 0156373
Lee, J. M. & Ullrich, A. (1984) Eur. Pat. Appl. 0128733
Liang, S.-M., Allet, B., Rose, K., Hirschi, M., Liang, C.-M. & Thatcher, D. R. (1985) Biochem. J. **229**, 429–439
Liu, F.-T., Albrandt, K. A., Bry, C. A. & Ishizaka, T. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 5369–5373
London, J., Skrzynia, C. & Goldberg, M. E. (1974) Eur. J. Biochem. **47**, 409–415
Lord, S. T. (1985) DNA **4**, 33–38
Lowe, P. A., Marston, F. A. O., Angal, S. & Schoemaker, J. A. (1984) World Pat. Appl. 8403711
McCaman, M. T., Andrews, W. H. & Files, J. G. (1985) J. Biotechnol **2**, 177
Malkovsky, M., Medawar, P. B., Thatcher, D. R., Toy, J., Hunt, R., Rayfield, L. S. & Dore, C. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 536–538
Mandecki, W., Mollison, K. W., Bollig, T. J., Powell, B. S., Carter, G. W. & Fox, J. L. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 3543–3547
Mark, D. F., Lin, L. S. & Yu-Lu, S.-D. (1983) Brit. U.K. Pat. Appl. 2130219
Marston, F. A. O., Lowe, P. A., Doel, M. T., Schoemaker, J. M., White, S. & Angal, S. (1984) Bio/Technology **2**, 800–804
Michaelis, S. & Beckwith, J. (1982) Annu. Rev. Microbiol. **36**, 435–465
Mozhaev, M. M. & Martinek, K. (1982) Enzyme Microb. Technol. **4**, 299–309
Muller, D., Hughes, C. & Goebel, W. (1983) J. Bacteriol. **153**, 846–851
Nagahari, K., Kanaya, S., Monakata, K., Aoyagi, Y. & Mizushima, S. (1985) EMBO J. **4**, 3589–3592
Nagai, K., Perutz, M. F. & Poyart, C. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7252–7255
Nicaud, J.-M., Mackman, N. & Holland, I. B. (1986) J. Biotechnol **3**, 255–270
Nishimori, K., Shimizu, N., Kawaguchi, Y., Hidaka, M., Uozumi, T. & Beppu, T. (1984) Gene **29**, 41–49
Nossal, N. G. & Heppel, L. A. (1966) J. Biol. Chem. **241**, 3055–3062
Offensperger, W., Wahl, S., Neurath, A. R., Price, P., Strick, N., Kent, S. B. H., Christman, J. K. & Acs, G. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7540–7544
Oka, T., Sakamoto, S., Miyoshi, K.-I., Fuwa, T., Yoda, K., Yamasaki, M., Tamura, G., Miyake, T. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7212–7216
Olsen, K. C., Fenno, J., Lin, N., Harkins, R. N., Snides, C., Kohr, W. H., Ross, M. J., Fodge, D., Prender, G. & Stebbing, N. (1981) Nature (London) **293**, 408–411
Olsen, K. C. (1985) U.S. Pat. 4518526
Olsen, K. C. & Pai, R.-C. (1985) U.S. Pat. 4511503
Pennica, D., Nedwin, G. E., Hayflick, J. S., Seeburg, P. H., Derynck, R., Palladino, M. A., Kohr, W. J., Aggarwal, B. B. & Goeddel D. V. (1984) Nature (London) **312**, 724–729
Pennica, D., Hayflick, J. S., Bringman, T. S., Palladina, M. A. & Goeddel, D. V. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 6060–6064
Pilancinski, W. P., Glassman, D. L., Krzyzek, R. A., Sadowski, P. L. & Robbins, A. K. (1984) Bio/Technology **2**, 356–360
Pollitt, S. & Zalkin, H. (1983) J. Bacteriol. **153**, 27–32
Pollock, M. R. & Richmond, M. H. (1962) Nature (London) **194**, 446–449
Prouty, W. F. & Goldberg, A. L. (1972) Nature (London) **240**, 147–150
Prouty, W. F., Karnovsky, M. J. & Goldberg, A. L. (1975) J. Biol. Chem. **250**, 1112–1122
Sassenfeld, H. M. & Brewer, S. J. (1984) Bio/Technology **2**, 76–81
Schoemaker, J. M., Brasnett, A. H. & Marston, F. A. O. (1985) EMBO J. **4**, 775–780
Schoner, R. G., Ellis, L. F. & Schoner, B. E. (1985) Bio/Technology, **3**, 151–154
Seeburg, P. H., Shine, J., Martial, J. A., Ivarie, R. D., Morris, J. E., Ulltrich, A., Baxter, J. D. & Goodman, H. M. (1978) Nature (London) **276**, 795–798
Sekine, S., Mizukami, T., Nishi, T., Kuwana, Y., Saito, A., Sato, M., Itoh, S. & Kawauchi, H. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4306–4310
Shen, S.-H. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 4627–4631
Shine, J., Fettes, I., Lan, N. C. Y., Roberts, J. L. & Baxter, J. D. (1980) Nature (London) **285**, 456–461
Silhavy, T. J., Benson, S. A. & Emr, S. D. (1983) Microbiol. Rev. **47**, 313–344
Simons, G., Remaut, E., Allet, B., Devos, R. & Fiers, W. (1984) Gene **28**, 55–64
Sofer, G. (1984) Bio/Technology **2**, 1035–1038
Staehelin, T., Hobbs, D. S., Kung, H.-F. & Pestka, S. (1981) Methods Enzymol. **78**, 505–511
Stebbing, N., Lee, S. H., Marcifino, B. J., Weck, P. K. & Renton, K. W. (1982) in From Gene to Protein Translation into Biotechnology (Ahmad, F., Smith, E. E., Schuttz, J. & Whelan, W. J., eds.), pp. 445–458, Academic Press, New York
Straus, D. & Gilbert, W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 2014–2018
Sung, W. L., Yao, F.-L., Zahab, D. M. & Narang, S. A. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 561–565
Szoka, P. R., Schreiber, A. B., Chan, H. & Murthy, J. (1986) DNA **5**, 11–20
Talmadge, K. & Gilbert, W. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 1830–1833
Tommassen, J., Van Tol, H. & Lugtenberg, B. (1983) EMBO J. **2**, 1275–1279
Varadarajan, R., Szabo, A. & Boxes, S. G. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 5681–5684
Villa-Komaroff, L., Erfstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) Proc. Natl. Acad. Sci. U.S.A. **75**, 3727–3731
Vogel, T., Weisgraer, K. H., Zeevi, M. I., Ben-Artzi, H., Levanon, A. Z., Rall, S. C., Jr., Innerarity, T. L., Hui, D. Y., Taylor, J. M., Kanner, D., Yavin, Z., Amit, B., Aviv, H., Gorecki, M. & Mahley, R. W. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 8696–8700
Waldman, A. S., Haeusslein, E. & Milman, G. (1983) J. Biol. Chem. **258**, 11571–11575
Wetzel, R. & Goeddel, D. V. (1983) Peptides **5**, 1–64
Wetzel, R., Heyneker, H. L., Goeddel, D. V., Juhrani, P., Shapiro, J., Crea, R., Low, T. L., McClure, K., Thurman, J. E. & Goldstein, A. L. (1981) in Cellular Responses to Molecular Modulators (Mozes, L. W., Schultz, J., Scott, W. A. & Werner, R., eds.), pp. 251–266, Academic Press, New York
Whitehorn, E. A., Livak, K. J. & Petteway, S. R., Jr. (1985) Gene **36**, 375–379
Williams, D. C., Van Frank, R. M., Muth, R. M. & Burnett, J. P. (1982) Science **215**, 687–689
Winkler, M. E., Blaber, M., Bennett, G. L., Holmes, W. & Vehar, G. A. (1985) Bio/Technology **3**, 990–1000
Zemel-Dreasen, O. & Zamir, A. (1984) Gene **27**, 315–322